PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/57, 9/64, A61K 31/70, 38/48, C12Q 1/37, A61K 48/00, C07K 16/40, C12N 15/00, A01K 67/027, C12Q 1/68

A1

(11) International Publication Number: **WO 98/49322**

(43) International Publication Date: 5 November 1998 (05.11.98)

(21) International Application Number: PCT/IB98/00625

(22) International Filing Date: 24 April 1998 (24.04.98)

(30) Priority Data:
966/97 26 April 1997 (26.04.97) CH

(71)(72) Applicant and Inventor: SONDEREGGER, Peter [CH/CH]; Biochemisches Institut Universität Zürich, Winterthurerstrasse 190, CH–8057 Zürich (CH).

(74) Agent: ZINK–WILD, Markus, Peter; Patentanwaltsbüro Zink, Birchlistrasse 11, CH–8173 Riedt–Neerach (CH).

(81) Designated States: AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GE, GW, HU, ID, IL, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NEUROTRYPSIN

(57) Abstract

There are described neurotrypsins of the formulas (I) or (II), including the separate coding and coded sequences of these compounds of the formulas (I) or (II). These compounds may be used as at least one active compound in a pharmaceutical component. The coded peptide sequences of these compounds may be used as targets for the development of pharmaceutical drugs.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Neurotrypsin

Technical Field

The present invention is directed to neurotrypsins and to a pharmaceutical composition which contains these substances or has an influence on these substances.

Disclosure of Invention

Neurotrypsin is a newly discovered serine protease, which is predominantly expressed in the brain and in the lungs; the expression in the brain takes place nearly exclusively in the neurons.

Neurotrypsin has a previously not yet found domain composition: besides the protease domain, there are found 3 or 4 SRCR (scavenger receptor cysteine-rich) domains and one Kringle domain. It is to be pointed out that the combination of Kringle and SRCR domains have not yet been found in proteins. At the amino terminus of the neurotrypsin protein there is a segment of more than 60 amino acids, which has an extremely high proportion of proline and basic amino acids (arginine and histidine).

The invention is characterized by the characteristics in the independent claims. Preferred embodiments are defined in the dependent claims.

The newly found neurotrypsins

- neurotrypsin of the human (compound of the formula I),
- neurotrypsin of the mouse (compound of the formula II)

differ structurally very much from the so far known serine proteases.

The serine protease whose protease domain is structurally most closely related with the protease domain of the new compounds, namely plasmin (of the human), has only a 44 % amino acid sequence identity.

Th proline-rich, basic segment at the amino terminus has a certain resemblance with th basic segments of the netrins and the semaphorins/collapsins. Due to this

segment, it is probable that neurotrypsin may be enriched by means of heparin-affinity chromatography.

The neurotrypsins of the human (compound of the formula I) and of the mouse (compound of the formula II) exhibit a very high structural similarity among each other.

The identity of the amino acid sequences of the native proteins of the compounds of the formulas I or II amounts to 81%.

The neurotrypsin of the human (compound of the formula I) has a coding sequence of 2625 nucleotides. The coded peptide of the compound of the formula I has a length of 875 amino acids and contains a signal peptide of 20 amino acids. The neurotrypsin of the mouse (compound of the formula II) has a coding sequence of 2283 nucleotides. The coded protein of the compound of the formula II has a length of 761 amino acids and contains a signal peptide of 21 amino acids. The reason for the greater length of the neurotrypsin of the human consists therein that the human neurotrypsin has 4 SRCR domains, whereas the neurotrypsin of the mouse has only 3 SRCR domains.

The domains which are present in both compounds (compound of the formula I and compound of the formula II) have a high degree of sequence similarity. The corresponding SRCR domains of the compounds of the formulas I and II have an amino acid sequence identity from 81% to 91%. The corresponding Kringle domains have an amino acid sequence identity of 75%. A high degree of similarity consists also in the enzymatically active (i.e. proteolytic) domain (90% amino acid sequence identity).

The protease domains of the neurotrypsins of the human (compound of the formula I) and of the mouse (compound of the formula II) are aligned in the following section, in order to illustrate the high degree of sequence identity.

```
CGLRLLHRRQKRIIGGKNSLRGGWPWQVSLRLKSSHGDGRLLCGATLLSS          50
|||||||||||||||||.|||||:||||.||||:|.||||||||||||||
CGLRLLHRRQKRIIGGNNSLRGAWPWQASLRLRSAHGDGRLLCGATLLSS

CWVLTAAHCFKRYGNSTRSYAVRVGDYHTLVPEEFEEEIGVQQIVIHREY          100
|||||||||||||||..|||||||||||||||||:||||||||||:|
CWVLTAAHCFKRYGNNSRSYAVRVGDYHTLVPEEFEQEIGVQQIVIHRNY

RPDRSDYDIALVRLQGPEEQCARFSSHVLPACLPLWRERPQKTASNCYIT          150
||||||||||||||||:|||||:|.|||||||||||||||||.||
RPDRSDYDIALVRLQGPGEQCARLSTHVLPACLPLWRERPQKTASNCHIT

GWGDTGRAYSRTLQQAAIPLLPKRFCEERYKGRFTGRMLCAGNLHEHKRV          200
||||||||||||||||:||||||||.|||||  ||||||||||:|..||
GWGDTGRAYSRTLQQAAVPLLPKRFCKERYKGLFTGRMLCAGNLQEDNRV

DSCQGDSGGPLMCERPGESWVVYGVTSWGYGCGVKDSPGVYTKVSAFVPW          250
|||||||||||||:|:||||||||||||||||||.|||||:|.|||||
DSCQGDSGGPLMCEKPDESWVVYGVTSWGYGCGVKDTPGVYTRVPAFVPW

IKSVTKL                                                     258
|||||.|
IKSVTSL
```

From the 258 amino acid sequence positions included in the comparison there are 233 amino acids that are identical in both compounds (upper sequence: compound of the formula I; lower sequence: compound of the formula II; identical amino acids are indicated by vertical lines).

The inventive neurotrypsins are unique when compared with the known serine proteases in that they are expressed according to currently available observations in a distinct degree in neurons. A further organ with a strong expression of neurotrypsin are the lungs (see Gschwend et al., Mol. Cell. Neurosci. 9, pages 207-219, 1997).

The proteins that are structurally most similar to the compounds of the formulas I or II are serine proteases, such as tissue-type plasminogen activator (tPA), urokinase-type plasminogen activator (uPA), plasmin, trypsin, apolipoprotein (a), coagulation factor XI, neuropsin, and acrosin.

In the adult brain, the inventive compounds are expressed predomiantly in the cerebral cortex, the hippocampus, and the amygdala.

In the adult brain stem and the spinal cord, the inventive compounds are expressed predominantly in the motor neurons. A slightly weaker expression is found in the neurons of the superficial layers of the dorsal horn of the spinal cord.

In the adult peripheral nervous system, the inventive compounds are expressed in a subpopulation of the sensory ganglia neurons.

The inventive compounds were found in connection with a study aimed at discovering trypsin-like serine proteases in the nervous system.

The first compound that was found and characterized was the compound of the formula II (Gschwend et al., Mol. Cell. Neurosci. 9, pages 207-219, 1997).

By means of an alignment of the protease domains of 7 known serine proteases (tissue-type plasminogen activator, urokinase-type plasminogen activator, thrombin, plasmin, trypsin, chymotrypsin, and pancreatic elastase) in the proximity of the histidine and the serine of the catalytic triade of the active site, the sequences of the so-called primer oligonucleotides for the polymerase chain reaction were determined.

The primer oligonucleotides were used in a polymerase chain reaction (PCR) together with ss-cDNA from total RNA of the brains of 10 days old mice and resulted in the amplification of a cDNA fragment of a length of approximately 500 base pairs.

This cDNA fragment was used successfully for the isolation of further cDNA fragments by screening commercially available cDNA libraries. Together, the isolated cDNA fragments covered the full length of the coding part of the compound of the formula II.

By conventional DNA sequencing the complete nucleotide sequence and the amino acid sequence deduced therefrom was obtained.

The compound of the formula I was cloned based on its pronounced similarity with the compound of the formula II.

The primer oligonucleotides used were synthesized according to the known sequence of the compound of the formula II.

The cloning of the compound of the formula I was performed by means of two commercially available cDNA libraries from fetal human brain.

This procedure for the cloning can also be used for the isolation of the homologous compounds of other species, such as rat, rabbit, guinea pig, cow, sheep, pig, primates, birds, zebra fish (Brachydanio rerio), Drosophila melanogaster, Caenorhabditis elegans etc.

The coding nucleotide sequences can be used for the production of proteins with the coded amino acid sequences of the compounds of the formulas I or II. A procedure developed in our laboratory allows the production of recombinant proteins in myeloma cells as fusion proteins with an immunoglobulin domain (constant domain of the kappa light chain). The principle of the construction is given in detail by Rader et al. (Rader et al., Eur. J. Biochem. 215, pages 133-141, 1993). The fusion protein produced by the myeloma cells was isolated by immunoaffinity chromatography using a monoclonal antibody against the Ig domain of the kappa light chain. With the same expression method, also the native protein of a compound, starting from the coding sequence, can be produced.

The coding sequences of the compounds of the formulas I or II can be used as starting compounds for the discovery and the isolation of alleles of the compounds of the formulas I or II. Both the polymerase chain reaction and the nucleic acid hybridization can be used for this purpose.

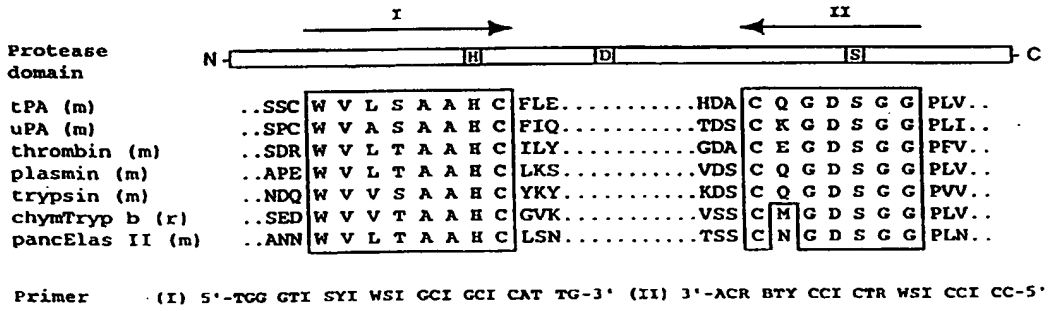
The coding sequences of the compounds of the formulas I or II can be used as starting compounds for so-called "site-directed mutagenesis", in order to generate nucleotide sequences coding the coded proteins that are defined by the compounds of the formulas I or II, or parts thereof, but whose nucleotide sequence is degenerated with respect to the compounds of the formulas I or II due to use of alternative codons.

The coding sequences of the compounds of the formulas I or II can be used as starting compounds for the production of sequence variants by means of so-called site-directed mutagenesis.

Best Mod s for Carrying out the Invention (Examples)

cDNA cloning of the c mpound of the f rmula II (neurotrypsin of the m use)

Total RNA was isolated from the brains of 10 days old mice (ICR-ZUR) according to the method of Chomczynski and Sacchi (1987). The production of single stranded cDNA was carried out using oligo(dT) primer and a RNA-dependent DNA polymerase (SuperScript RNase H-Reverse Transcriptase; Gibco BRL, Gaithersburg, MD) according to the instruction of the supplier. For the realization of the polymerase chain reaction one forward primer was synthesized based on the amino acid sequence of the region of the conserved histidine of the catalytic triade and one primer in the backward direction was synthesized based on the amino acid sequence of the region of the conserved serine of the catalytic triade of the serine proteases. The amino acid sequences used for the determination of the oligonucleotide primers were taken from seven known serine proteases. They are presented in the following.



The protease domains of 7 known serine proteases (tissue-type plasminogen activator, urokinase-type plasminogen activator, thrombin, plasmin, trypsin, chymotrypsin, and pancreatic elastase) were aligned in the region of the conserved histidine and serine of the catalytic triade of the active site. The conserved amino acids of these regions were taken as the basis for the determination of the degenerated primers. The primer sequences are given according to the recommendation of the IUB nomenclature (Nomenclature Committee 1985).

The primers used in the PCR contained restriction sites for *Eco*RI and *Bam*HI at their 5' ends in order to facilitate a subsequent cloning.

The following primers were used:

In the reading direction (sense primers):

5'-GGGGAATTCTGGGTI(C/G)(T/C)I(T/A)(G/C)IGCIGCICA(T/C)TG-3'

In the counter direction (antisense primers):

5'-GGGGGATCCCCICCI(G/C)(A/T)(A/G)TCICC(C/T)T(G/C/T)(G/A)CA-3'.

The polymerase chain reaction was carried out under standard conditions using the DNA polymerase AmpliTaq (Perkin Elmer) according to the recommendations of the producer. The following PCR profile was employed: 93°C for 3 minutes, followed by 35 cycles of 93°C for 1 minute, 48°C for 2 minutes, and 72°C for 2 minutes. Following the last cycle, the incubation was continued at 72°C for further 10 minutes.

The amplified fragments had an approximate length of 500 base pairs. They were cut with *Eco*RI and *Bam*HI and inserted in a Blue Script vector (Bluescript SK(-), Stratagene). The resulting clones were analyzed by DNA sequence determination using the dideoxy chain termination method (Sanger et al., Proc. Natl. Acad. Sci. USA 77, pages 2163-2167, 1977) on an automated DNA sequencer (LI-COR, model 4000L; Lincoln, NE) using a commercial sequencing kit (SequiTerm long-read cycle sequencing kit-LC; Epicentre Technologies, Madison, WI). The analysis yielded a sequence of 474 base pairs of the catalytic region of the serine protease domain of the compound of the formula II.

The 474 base pair long PCR fragment was used for screening of an oligo(dT)-primed Uni-ZAP-XR cDNA library from the brain of 20 days old mice (Stratagene; cat. no. 937 319). At total of 3 x 10^6 lambda plaques were screened under high stringent conditions (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989) using a radioactively labeled PCR fragment as a probe and 24 positive clones were found.

From the positive Lambda-Uni-ZAP-XR phagemid clones the corresponding Bluescript plasmid was cut out by *in vivo* excision according to a standard method recommended by the producer (Stratagene). In order to determine the length of the inserted fragm nts the corresponding Bluescript plasmid clones were digested with *Sac*I and *Kpn*I. The clones containing the longest fragments wer analyzed by DNA

sequencing (as described above) and for subsequent data analysis the GCG software (version 8.1, Unix; Silicon Graphics, Inc.) was used.

Because none of the clones contained the coding sequence in full length, a second cDNA library was screened. The library used in this screen was an oligo(dT)- and random-primed cDNA library in a Lambda phage (Lambda gt10) which was based on mRNA from 15 days old mouse embryos (oligo(dT)- and random-primed Lambda gt10 cDNA library; Clontech, Palo Alto, CA; cat. no. ML 3002a). As a probe a radioactively labeled DNA fragment (AvaI/AatII) from the 5' end of the longest clone of the first screen was used and approximately $2x10^6$ plaques were screened. This screen resulted in 14 positive clones. The cDNA fragments were excised with *Eco*RI and cloned into the Bluecript vector (KS(+); Stratagene). The sequence analysis was carried out as described above.

In this way the nucleotide sequence over the full length cDNA of 2361 and 2376 base pairs, respectively, of the compound of the formula II was obtained. With the described procedure of PCR cloning it is possible to find and isolate also variant forms of the compounds of the formulas I or II, as for example their alleles or their splice variants. The described method of screening of a cDNA library allows also the detection and the isolation of compounds which hybridize under stringent conditions with the coding sequences of the compounds of the formulas I or II.

Cl ning of the cDNA of th compound f the formula I (neurotrypsin of the human)

The cloning of the cDNA of the compound of the formula I was carried out basing on the nucleotide sequence of the compound of the formula II. As a first step, a fragment of the compound of the formula I was amplified using the polymerase chain reaction (PCR). As a matrix we used the DNA obtained from a cDNA library from the brain of a human fetus (17th - 18th week of pregnancy) which is commercially available (Oligo(dT)- and random-primed, human fetal brain cDNA library in the Lambda ZAP II vector, cat. no. 936206, Stratagene). The synthetic PCR primers contained restriction sites for *Hind*III and *Xho*I at the 5' end in order to facilitate the subsequent cloning.

In the reading direction (sense primers):

5'-GGGAAGCTTGGICA(A/G)TGGGGIACI(A/G)TITG(C/T)GA(C/T)-3'

In the counter direction (antisense primers):

5'-GGGCTCGAGCCCCAICCTGTTATGTAAIAGTTG-3'

The PCR was carried out under standard conditions using the DNA polymerase Amplitaq (Perkin Elmer) according to the recommendations of the producer. The resulting fragment of 1116 base pairs was inserted into the Bluescript vector (Bluescript SK(-), Stratagene). A 600 base pairs long *Hind*III/*Stu*I fragment, corresponding to the 5' half the 1116 base pairs long PCR fragment, was used for the screening of a Lamda cDNA library from human fetal brain (Human Fetal Brain 5'-STRETCH PLUS cDNA library; Lambda gt10; cat. no. HL 3003 a; Clontech). $2x10^6$ Lambda plaques were screened under high stringent conditions (Sambrook et al., Molecular Cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, 1989) by means of a radioactively labeled PCR fragment, and 23 positive clones were found and isolated.

From the positive Lambda gt10 clones the corresponding cDNA fragments were excised with *Eco*RI and inserted into a Bluescript vector (Bluescript KS(+), Stratagene). The sequencing was carried out by means of the dideoxy chain termination method (Sanger et al., Proc. Natl. Acad. Sci. USA 77, pages 2163-2167, 1977), using a commercial sequencing kit (SequiTherm long-read cycle sequencing kit-LC; Epicentre Technologies, Madison, WI) and Bluescript-specific primers.

In an alternative sequencing strategy, the cDNA fragments of th positive Lambda gt10 clones were PCR amplified using Lambda-specific primers. The sequencing was carried out as described above.

The computerized analysis of the sequences was performed by means of the program package GCG (version 8.1, Unix; Silicon Graphics Inc.).

In this way the nucleotide sequence over the full length of the cDNA of 3350 base pairs was obtained. With the described procedure for PCR cloning it is possible to find and to isolate also variant forms of the compounds of the formulas I or II, as for example their alleles or their splice variants. The described procedure for the screening of a cDNA library allows also the discovery and the isolation of compounds which hybridize under stringent conditions with the coding sequences of the compounds of the formulas I or II.

Visualization of the coded sequences of the compounds of the formulas I or II by means of antibodies

The more than 60 amino acids long proline-rich, basic segment at the amino terminus of the coded sequence of the compounds of the formulas I or II is well suited for the production of antibodies by means of synthesizing peptides and using them for immunization. We have selected two peptide sequences with a length of 19 and 13 amino acids from the proline-rich, basic segment at the amino terminus of the coded sequence of the compound of the formula II for the generation of antibodies. The peptides had the following sequences:

Peptide 1: H_2N-SRS PLH RPH PSP PRS QX-$CONH_2$

Peptide 2: H_2N-LPS SRR PPR TPR F-COOH

The two peptides were synthesized chemically, coupled to a macromolecular carrier (Keyhole Limpet Hemacyanin), and injected into 2 rabbits for immunization. The resulting antisera exhibit a high antibody titer and could successfully be used both for the identification of native neurotrypsin in brain extract of the mouse and for the identification of recombinant neurotrypsin. The employed procedure for the generation of antibodies can also be used for the generation of antibodies against the coded sequence of the compound of the formula I.

The resulting antibodies against the partial sequences of the coded sequences of the compounds of the formulas I or II can be used for the detection and the isolation of variant forms of the compounds of the formulas I or II, as for example alleles or splice variants. Such antibodies can also be used for the detection and isolation of gene technologically generated variants of the compounds of the formulas I or II.

Purification of the coded sequences of the compounds of the formulas I or II

Besides conventional chromatographic methods, as for example ion exchange chromatography, the purification of the coded sequences of the compounds of the formulas I or II can also be achieved using two affinity chromatographic purification procedures. One affinity chromatographic purification procedure is based on the availability of antibodies. By coupling the antibodies on a chromatographic matrix, a purification procedure results, in which a very high degree of purity of the corresponding compound can be achieved in one step.

Another important feature that can be used for the purification of the coded sequences of the compounds of the formulas I or II is the proline-rich, basic segment at the amino terminus. It may be expected that, due to the high density of positive charges, this segment mediates the binding of the coded sequences of the compounds of the formulas I or II to heparin and heparin-like affinity matrices. This principle allows also the isolation, or at least the enrichment, of variant forms of the coded sequences of the compounds of the formulas I or II, as for example their alleles or splice variants. Likewise the heparin affinity chromatography can be used for the isolation, or at least the enrichment, of species-homologous proteins of the compounds of the formulas I or II.

Industrial Applicability

The coding sequences of the formulas I and II can be used for the production of the coded proteins or parts thereof of the formulas I and II. The production of the coded proteins can be achieved in procaryotic or eucaryotic expression systems.

The gene expression pattern of the inventive compounds in the brain is extremely interesting, because these molecules are expressed in the adult nervous system predominantly in neurons of those regions that are thought to play an important role in learning and memory functions. Together with the recently found evidence for a role of extracellular proteases in neural plasticity, the expression pattern allows the assumption that the proteolytic activity of neurotrypsin has a role in structural reorganizations in connection with learning and memory operations, for example operations which are involved in the processing and storage of learned behaviors, learned emotions, or memory contents. The inventive compounds may, thus, represent a target for pharmaceutical intervention in malfunctions of the brain.

The gene expression pattern of the inventive compounds in the cerebral cortex (especially layers V and VI) is extremely interesting, because a reduction of the cellular differentiation in the cerebral cortex has been found to be associated with schizophrenia. The inventive compounds may, thus, be a target for pharmaceutical intervention in schizophrenia and related psychiatric diseases.

The coding sequences of the inventive compounds have been found to be increased in the neurons located adjacent to the damaged tissue of a focal ischemic stroke, indicating that the inventive compounds play a role in the tissue reaction in the injured cerebral tissue. The inventive compounds may, thus, represent a target for pharmaceutical intervention after ischemic stroke and other forms of neural tissue damage.

Tissue-type plasminogen activator, a serine protease related to the inventive compounds, has recently been found to be involved in excitotoxicity-mediated neuronal cell death. A similar function is conceivable for the inventive compounds and, thus, the inventive compounds represent a possible target for a pharmacological intervention in diseases in which cell death occurs.

The gene expression pattern of the inventive compounds in the spinal cord and in the sensory ganglia is interesting, because these molecules are expressed in the adult nervous system in neurons of those brain regions that are thought to play a role in the processing of pain, as well as in the pathogenesis of pathological pain. The inventive compounds may, thus, be a target for pharmaceutical intervention in pathological pain.

In the following part statements concerning the compounds of the formulas I or II are given:

(1) INFORMATION ABOUT THE COMPOUND OF THE FORMULA I
(Neurotrypsin of the human)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3350 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single strand
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens
(D) DEVELOPMENT STAGE: fetal
(F) TISSUE TYPE: brain

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: human fetal brain 5'-stretch plus cDNA library in the lambda gt10 vector; catalog No. HL 3003a; Clontech, Palo Alto, CA, USA.

(B) CLONE: cDNA Clone No.:
3-1, 3-2, 3-6, 3-7, 3-8, 3-10, 3-11, 3-12

(ix) FEATURE:

(A) NAME/KEY: Signal peptide
(B) LOCATION: 44 .. 103

(ix) FEATURE:

(A) NAME/KEY: mature peptide
(B) LOCATION: 104 .. 2668

(ix) FEATURE:

(A) NAME/KEY: coding sequence
(B) LOCATION: 44 .. 2668

(ix) FEATURE:

(A) NAME/KEY: Proline-rich, basic segment
(B) LOCATION: 104 .. 319

(ix) FEATURE:

(A) NAME/KEY: Kringle domain
(B) LOCATION: 320 .. 538

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 1
(B) LOCATION: 551 .. 856

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 2
(B) LOCATION: 881 .. 1186

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 3
(B) LOCATION: 1202 .. 1504

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 4
(B) LOCATION: 1541 .. 1846

(ix) FEATURE:

(A) NAME/KEY: proteolytic domain
(B) LOCATION: 1898 .. 2668

(ix) FEATURE:

(A) NAME/KEY: histidine of the catalytic triade
(B) LOCATION: 2069 - 2071

(ix) FEATURE:

(A) NAME/KEY: aspartic acid of the catalytic triade
(B) LOCATION: 2219 - 2221

(ix) FEATURE:

(A) NAME/KEY: serine of the catalytic triade
(B) LOCATION: 2516 .. 2518

(ix) FEATURE:

(A) NAME/KEY: polyA signal
(B) LOCATION: 2873 .. 2878

(ix) FEATURE

(A) NAME/KEY: polyA signal
(B) LOCATION: 3034 .. 3039

(ix) FEATURE:

(A) NAME/KEY: polyA signal
(B) LOCATION: 3215 .. 3220

(ix) FEATURE:

(A) NAME/KEY: 3'UTR
(B) LOCATION: 2669 .. 3350

(ix) FEATURE

(A) NAME/KEY: 5'UTR
(B) LOCATION: 1 .. 43

C mp und of th formula I (neur trypsin f the human)

```
CGGAAGCTGG GGAGCATGGA CCAGACCCCG CAGCGCTGGC ACC ATG ACG CTC GCC      55
                                                 Met Thr Leu Ala
                                                 -20

CGC TTC GTG CTA GCC CTG ATG TTA GGG GCG CTC CCC GAA GTG GTC GGC     103
Arg Phe Val Leu Ala Leu Met Leu Gly Ala Leu Pro Glu Val Val Gly
    -15                 -10                 -5                 -1

TTT GAT TCT GTC CTC AAT GAT TCC CTC CAC CAC AGC CAC CGC CAT TCG     151
Phe Asp Ser Val Leu Asn Asp Ser Leu His His Ser His Arg His Ser
  1               5                  10                  15

CCC CCT GCG GGT CCG CAC TAC CCC TAT TAC CTT CCC ACC CAG CAG CGG     199
Pro Pro Ala Gly Pro His Tyr Pro Tyr Tyr Leu Pro Thr Gln Gln Arg
            20                  25                  30

CCC CCG ACG ACG CGT CCG CCG CCG CCT CTC CCG CGC TTC CCG CGC CCC     247
Pro Pro Thr Thr Arg Pro Pro Pro Pro Leu Pro Arg Phe Pro Arg Pro
        35                  40                  45

CCG CGG GCG CTC CCT GCC CAG CGC CCG CAC GCC CTC CAG GCC GGG CAC     295
Pro Arg Ala Leu Pro Ala Gln Arg Pro His Ala Leu Gln Ala Gly His
    50                  55                  60

ACG CCC CGG CCG CAC CCC TGG GGC TGC CCC GCC GGC GAG CCA TGG GTC     343
Thr Pro Arg Pro His Pro Trp Gly Cys Pro Ala Gly Glu Pro Trp Val
65                  70                  75                  80

AGC GTG ACG GAC TTC GGC GCC CCG TGT CTG CGG TGG GCG GAG GTG CCA     391
Ser Val Thr Asp Phe Gly Ala Pro Cys Leu Arg Trp Ala Glu Val Pro
                85                  90                  95

CCC TTC CTG GAG CGG TCG CCC CCA GCG AGC TGG GCT CAG CTG CGA GGA     439
Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala Gln Leu Arg Gly
            100                 105                 110

CAG CGC CAC AAC TTT TGT CGG AGC CCC GAC GGC GCG GGC AGA CCC TGG     487
Gln Arg His Asn Phe Cys Arg Ser Pro Asp Gly Ala Gly Arg Pro Trp
        115                 120                 125

TGT TTC TAC GGA GAC GCC CGT GGC AAG GTG GAC TGG GGC TAC TGC GAC     535
Cys Phe Tyr Gly Asp Ala Arg Gly Lys Val Asp Trp Gly Tyr Cys Asp
    130                 135                 140

TGC AGA CAC GGA TCA GTA CGA CTT CGT GGC GGC AAA AAT GAG TTT GAA     583
Cys Arg His Gly Ser Val Arg Leu Arg Gly Gly Lys Asn Glu Phe Glu
145                 150                 155                 160

GGC ACA GTG GAA GTA TAT GCA AGT GGA GTT TGG GGC ACT GTC TGT AGC     631
Gly Thr Val Glu Val Tyr Ala Ser Gly Val Trp Gly Thr Val Cys Ser
                165                 170                 175

AGC CAC TGG GAT GAT TCT GAT GCA TCA GTC ATT TGT CAC CAG CTG CAG     679
Ser His Trp Asp Asp Ser Asp Ala Ser Val Ile Cys His Gln Leu Gln
            180                 185                 190
```

```
CTG GGA GGA AAA GGA ATA GCA AAA CAA ACC CCG TTT TCT GGA CTG GGC      727
Leu Gly Gly Lys Gly Ile Ala Lys Gln Thr Pro Phe Ser Gly Leu Gly
        195                 200                 205

CTT ATT CCC ATT TAT TGG AGC AAT GTC CGT TGC CGA GGA GAT GAA GAA      775
Leu Ile Pro Ile Tyr Trp Ser Asn Val Arg Cys Arg Gly Asp Glu Glu
    210                 215                 220

AAT ATA CTG CTT TGT GAA AAA GAC ATC TGG CAG GGT GGG GTG TGT CCT      823
Asn Ile Leu Leu Cys Glu Lys Asp Ile Trp Gln Gly Gly Val Cys Pro
225                 230                 235                 240

CAG AAG ATG GCA GCT GCT GTC ACG TGT AGC TTT TCC CAT GGC CCA ACG      871
Gln Lys Met Ala Ala Ala Val Thr Cys Ser Phe Ser His Gly Pro Thr
                245                 250                 255

TTC CCC ATC ATT CGC CTT GCT GGA GGC AGC AGT GTG CAT GAA GGC CGG      919
Phe Pro Ile Ile Arg Leu Ala Gly Gly Ser Ser Val His Glu Gly Arg
            260                 265                 270

GTG GAG CTC TAC CAT GCT GGC CAG TGG GGA ACC GTT TGT GAT GAC CAA      967
Val Glu Leu Tyr His Ala Gly Gln Trp Gly Thr Val Cys Asp Asp Gln
        275                 280                 285

TGG GAT GAT GCC GAT GCA GAA GTG ATC TGC AGG CAG CTG GGC CTC AGT     1015
Trp Asp Asp Ala Asp Ala Glu Val Ile Cys Arg Gln Leu Gly Leu Ser
    290                 295                 300

GGC ATT GCC AAA GCA TGG CAT CAG GCA TAT TTT GGG GAA GGG TCT GGC     1063
Gly Ile Ala Lys Ala Trp His Gln Ala Tyr Phe Gly Glu Gly Ser Gly
305                 310                 315                 320

CCA GTT ATG TTG GAT GAA GTA CGC TGC ACT GGG AAT GAG CTT TCA ATT     1111
Pro Val Met Leu Asp Glu Val Arg Cys Thr Gly Asn Glu Leu Ser Ile
                325                 330                 335

GAG CAG TGT CCA AAG AGC TCC TGG GGA GAG CAT AAC TGT GGC CAT AAA     1159
Glu Gln Cys Pro Lys Ser Ser Trp Gly Glu His Asn Cys Gly His Lys
            340                 345                 350

GAA GAT GCT GGA GTG TCC TGT ACC CCT CTA ACA GAT GGG GTC ATC AGA     1207
Glu Asp Ala Gly Val Ser Cys Thr Pro Leu Thr Asp Gly Val Ile Arg
        355                 360                 365

CTT GCA GGT GGG AAA GGC AGC CAT GAG GGT CGC TTG GAG GTA TAT TAC     1255
Leu Ala Gly Gly Lys Gly Ser His Glu Gly Arg Leu Glu Val Tyr Tyr
    370                 375                 380

AGA GGC CAG TGG GGA ACT GTC TGT GAT GAT GGC TGG ACT GAG CTG AAT     1303
Arg Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp Thr Glu Leu Asn
385                 390                 395                 400

ACA TAC GTG GTT TGT CGA CAG TTG GGA TTT AAA TAT GGT AAA CAA GCA     1351
Thr Tyr Val Val Cys Arg Gln Leu Gly Phe Lys Tyr Gly Lys Gln Ala
                405                 410                 415

TCT GCC AAC CAT TTT GAA GAA AGC ACA GGG CCC ATA TGG TTG GAT GAC     1399
Ser Ala Asn His Phe Glu Glu Ser Thr Gly Pro Ile Trp Leu Asp Asp
            420                 425                 430
```

```
GTC AGC TGC TCA GGA AAG GAA ACC AGA TTT CTT CAG TGT TCC AGG CGA    1447
Val Ser Cys Ser Gly Lys Glu Thr Arg Phe Leu Gln Cys Ser Arg Arg
        435                 440                 445

CAG TGG GGA AGG CAT GAC TGC AGC CAC CGC GAA GAT GTT AGC ATT GCC    1495
Gln Trp Gly Arg His Asp Cys Ser His Arg Glu Asp Val Ser Ile Ala
    450                 455                 460

TGC TAC CCT GGC GGC GAG GGA CAC AGG CTC TCT CTG GGT TTT CCT GTC    1543
Cys Tyr Pro Gly Gly Glu Gly His Arg Leu Ser Leu Gly Phe Pro Val
465                 470                 475                 480

AGA CTG ATG GAT GGA GAA AAT AAG AAA GAA GGA CGA GTG GAG GTT TTT    1591
Arg Leu Met Asp Gly Glu Asn Lys Lys Glu Gly Arg Val Glu Val Phe
                485                 490                 495

ATC AAT GGC CAG TGG GGA ACA ATC TGT GAT GAT GGA TGG ACT GAT AAG    1639
Ile Asn Gly Gln Trp Gly Thr Ile Cys Asp Asp Gly Trp Thr Asp Lys
            500                 505                 510

GAT GCA GCT GTG ATC TGT CGT CAG CTT GGC TAC AAG GGT CCT GCC AGA    1687
Asp Ala Ala Val Ile Cys Arg Gln Leu Gly Tyr Lys Gly Pro Ala Arg
        515                 520                 525

GCA AGA ACC ATG GCT TAC TTT GGA GAA GGA AAA GGA CCC ATC CAT GTG    1735
Ala Arg Thr Met Ala Tyr Phe Gly Glu Gly Lys Gly Pro Ile His Val
    530                 535                 540

GAT AAT GTG AAG TGC ACA GGA AAT GAG AGG TCC TTG GCT GAC TGT ATC    1783
Asp Asn Val Lys Cys Thr Gly Asn Glu Arg Ser Leu Ala Asp Cys Ile
545                 550                 555                 560

AAG CAA GAT ATT GGA AGA CAC AAC TGC CGC CAC AGT GAA GAT GCA GGA    1831
Lys Gln Asp Ile Gly Arg His Asn Cys Arg His Ser Glu Asp Ala Gly
                565                 570                 575

GTT ATT TGT GAT TAT TTT GGC AAG AAG GCC TCA GGT AAC AGT AAT AAA    1879
Val Ile Cys Asp Tyr Phe Gly Lys Lys Ala Ser Gly Asn Ser Asn Lys
            580                 585                 590

GAG TCC CTC TCA TCT GTT TGT GGC TTG AGA TTA CTG CAC CGT CGG CAG    1927
Glu Ser Leu Ser Ser Val Cys Gly Leu Arg Leu Leu His Arg Arg Gln
        595                 600                 605

AAG CGG ATC ATT GGT GGG AAA AAT TCT TTA AGG GGT GGT TGG CCT TGG    1975
Lys Arg Ile Ile Gly Gly Lys Asn Ser Leu Arg Gly Gly Trp Pro Trp
    610                 615                 620

CAG GTT TCC CTC CGG CTG AAG TCA TCC CAT GGA GAT GGC AGG CTC CTC    2023
Gln Val Ser Leu Arg Leu Lys Ser Ser His Gly Asp Gly Arg Leu Leu
625                 630                 635                 640

TGC GGG GCT ACG CTC CTG AGT AGC TGC TGG GTC CTC ACA GCA GCA CAC    2071
Cys Gly Ala Thr Leu Leu Ser Ser Cys Trp Val Leu Thr Ala Ala His
                645                 650                 655

TGT TTC AAG AGG TAT GGC AAC AGC ACT AGG AGC TAT GCT GTT AGG GTT    2119
Cys Phe Lys Arg Tyr Gly Asn Ser Thr Arg Ser Tyr Ala Val Arg Val
            660                 665                 670
```

```
GGA GAT TAT CAT ACT CTG GTA CCA GAG GAG TTT GAG GAA GAA ATT GGA    2167
Gly Asp Tyr His Thr Leu Val Pro Glu Glu Phe Glu Glu Glu Ile Gly
        675                 680                 685

GTT CAA CAG ATT GTG ATT CAT CGG GAG TAT CGA CCC GAC CGC AGT GAT    2215
Val Gln Gln Ile Val Ile His Arg Glu Tyr Arg Pro Asp Arg Ser Asp
    690                 695                 700

TAT GAC ATA GCC CTG GTT AGA TTA CAA GGA CCA GAA GAG CAA TGT GCC    2263
Tyr Asp Ile Ala Leu Val Arg Leu Gln Gly Pro Glu Glu Gln Cys Ala
705                 710                 715                 720

AGA TTC AGC AGC CAT GTT TTG CCA GCC TGT TTA CCA CTC TGG AGA GAG    2311
Arg Phe Ser Ser His Val Leu Pro Ala Cys Leu Pro Leu Trp Arg Glu
                725                 730                 735

AGG CCA CAG AAA ACA GCA TCC AAC TGT TAC ATA ACA GGA TGG GGT GAC    2359
Arg Pro Gln Lys Thr Ala Ser Asn Cys Tyr Ile Thr Gly Trp Gly Asp
            740                 745                 750

ACA GGA CGA GCC TAT TCA AGA ACA CTA CAA CAA GCA GCC ATT CCC TTA    2407
Thr Gly Arg Ala Tyr Ser Arg Thr Leu Gln Gln Ala Ala Ile Pro Leu
        755                 760                 765

CTT CCT AAA AGG TTT TGT GAA GAA CGT TAT AAG GGT CGG TTT ACA GGG    2455
Leu Pro Lys Arg Phe Cys Glu Glu Arg Tyr Lys Gly Arg Phe Thr Gly
    770                 775                 780

AGA ATG CTT TGT GCT GGA AAC CTC CAT GAA CAC AAA CGC GTG GAC AGC    2503
Arg Met Leu Cys Ala Gly Asn Leu His Glu His Lys Arg Val Asp Ser
785                 790                 795                 800

TGC CAG GGA GAC AGC GGA GGA CCA CTC ATG TGT GAA CGG CCC GGA GAG    2551
Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Glu Arg Pro Gly Glu
                805                 810                 815

AGC TGG GTG GTG TAT GGG GTG ACC TCC TGG GGG TAT GGC TGT GGA GTC    2599
Ser Trp Val Val Tyr Gly Val Thr Ser Trp Gly Tyr Gly Cys Gly Val
            820                 825                 830

AAG GAT TCT CCT GGT GTT TAT ACC AAA GTC TCA GCC TTT GTA CCT TGG    2647
Lys Asp Ser Pro Gly Val Tyr Thr Lys Val Ser Ala Phe Val Pro Trp
        835                 840                 845

ATA AAA AGT GTC ACC AAA CTG TAA TTCTTCATGG AAACTTCAAA GCAGCATTT    2700
Ile Lys Ser Val Thr Lys Leu  *
    850                 855

AAACAAATGG AAAACTTTGA ACCCCCACTA TTAGCACTCA GCAGAGATGA CAACAAATGG  2760

CAAGATCTGT TTTTGCTTTG TGTTGTGGTA AAAAATTGTG TACCCCCTGC TGCTTTTGAG  2820

AAATTTGTGA ACATTTTCAG AGGCCTCAGT GTAGTGGAAG TGATAATCCT TAAATGAACA  2880

TTTTCTACCC TAATTTCACT GGAGTGACTT ATTCTAAGCC TCATCTATCC CCTACCTATT  2940
```

TCTCAAAATC ATTCTATGCT GATTTTACAA AAGATCATTT TTACATTTGA ACTGAGAACC 3000

CCTTTTAATT GAATCAGTGG TGTCTGAAAT CATATTAAAT ACCCACATTT GACATAAATG 3060

CGGTACCCTT TACTACACTC ATGAGTGGCA TATTTATGCT TAGGTCTTTT CAAAAGACTT 3120

GACAAGAAAT CTTCATATTC TCTGTAGCCT TTGTCAAGTG AGGAAATCAG TGGTTAAAGA 3180

ATTCCACTAT AAACTTTTAG GCCTGAATAG GAGTAGTAAA GCCTCAAGGA CATCTGCCTG 3240

TCACAATATA TTCTCAAAGT GATCTGATAT TTGGAAACAA GTATCCTTGT TGAGTACCAA 3300

GTGCTACAGA AACCATAAGA TAAAAATACT TTCTACCTAC AGCGTGCCCG 3350

(1) INFORMATION ABOUT THE COMPOUND OF THE FORMULA II (Neurotrypsin of the mouse)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2376 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single strand
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mus musculus
(D) DEVELOPMENT STAGE: postnatal day 10
(F) TISSUE TYPE: brain
(G) CELL TYPE: neurons

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: mouse brain cDNA library in the lambda Uni-ZAP-XR vector, oligo (dT)-primed, from Balb c mice, postnatal day 20, Cat. No.. 937 319; Stratagene, La Jolla, CA, USA

(B) CLONE: cDNA clone no. 16

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: mouse brain cDNA library in the Lambda gt10 vector, oligo(dT)- and random-primed, embryonic day 15, Cat. No. ML 3002a; Clontech, Palo Alto, CA, USA

(B) CLONE: cDNA clone #25

(ix) FEATURE:

(A) NAME/KEY: signal peptide
(B) LOCATION: 24 .. 86

(ix) FEATURE:

(A) NAME/KEY: mature peptide
(B) LOCATION: 87 .. 2306

(ix) FEATURE:

(A) NAME/KEY: coding sequence
(B) LOCATION: 24 .. 2306

(ix) FEATURE:

(A) NAME/KEY: proline-rich, basic segment
(B) LOCATION: 90 .. 275

(ix) FEATURE:

(A) NAME/KEY: Kringle domain
(B) LOCATION: 276 .. 494

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 1
(B) LOCATION: 519 .. 824

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 2
(B) LOCATION: 840 .. 1142

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 3
(B) LOCATION: 1179 .. 1484

(ix) FEATURE:

(A) NAME/KEY: proteolytic domain
(B) LOCATION: 1536 .. 2306

(ix) FEATURE:

(A) NAME/KEY: histidine of the catalytic triade
(B) LOCATION: 1707 .. 1709

(ix) FEATURE:

(A) NAME/KEY: aspartic acid of the catalytic triade
(B) LOCATION: 1857 .. 1859

(ix) FEATURE:

(A) NAME/KEY: serine of the catalytic triade

(B) LOCATION: 2154 .. 2156

(ix) FEATURE:

(A) NAME/KEY:polyA signal
(B) LOCATION: 2324 .. 2329 and 2331 .. 2336

(ix) FEATURE:

(A) NAME/KEY: polyA segment
(B) LOCATION: 2357 .. 2376

(ix) FEATURE:

(A) NAME/KEY: 3'UTR
(B) LOCATION: 2307 .. 2341 or 2307 .. 2356

(ix) FEATURE:

(A) NAME/KEY: 5'UTR
(B) LOCATION: 1 .. 23

Coump und of the formula II (n ur trypsin of the mouse)

```
GGACCACACT CGGCGCCGCA GCC ATG GCG CTC GCC CGC TGC GTG CTG GCT GTG     53
                          Met Ala Leu Ala Arg Cys Val Leu Ala Val
                              -20                 -15

ATT TTA GGG GCA CTG TCT GTA GTG GCC CGC GCT GAT CCG GTC TCG CGC      101
Ile Leu Gly Ala Leu Ser Val Val Ala Arg Ala Asp Pro Val Ser Arg
    -10                 -5                  1               5

TCT CCC CTT CAC CGC CCG CAT CCG TCC CCA CCG CGT TCC CAA CAC GCG      149
Ser Pro Leu His Arg Pro His Pro Ser Pro Pro Arg Ser Gln His Ala
            10                  15                  20

CAC TAC CTT CCC AGC TCG CGG CGG CCA CCC AGG ACC CCG CGC TTC CCG      197
His Tyr Leu Pro Ser Ser Arg Arg Pro Pro Arg Thr Pro Arg Phe Pro
        25                  30                  35

CTC CCG CTG CGG ATC CCC GCT GCC CAG CGC CCG CAG GTC CTC AGC ACC      245
Leu Pro Leu Arg Ile Pro Ala Ala Gln Arg Pro Gln Val Leu Ser Thr
    40                  45                  50

GGG CAC ACG CCC CCG ACG ATT CCA CGC CGC TGC GGG GCA GGA GAG TCG      293
Gly His Thr Pro Pro Thr Ile Pro Arg Arg Cys Gly Ala Gly Glu Ser
55                  60                  65

TGG GGC AAT GCC ACC AAC CTC GGC GTC CCG TGT CTA CAC TGG GAC GAG      341
Trp Gly Asn Ala Thr Asn Leu Gly Val Pro Cys Leu His Trp Asp Glu
70                  75                  80                  85

GTG CCG CCC TTC CTG GAG CGG TCG CCC CCG GCC AGT TGG GCT GAG CTG      389
Val Pro Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala Glu Leu
                90                  95                  100

CGA GGG CAG CCG CAC AAC TTC TGC CGG AGC CCG GAT GGC TCG GGC AGA      437
Arg Gly Gln Pro His Asn Phe Cys Arg Ser Pro Asp Gly Ser Gly Arg
            105                 110                 115

CCT TGG TGC TTC TAT CGG AAT GCC CAG GGC AAA GTA GAC TGG GGC TAC      485
Pro Trp Cys Phe Tyr Arg Asn Ala Gln Gly Lys Val Asp Trp Gly Tyr
        120                 125                 130

TGC GAT TGT GGT CAA GGC CCG GCG TTG CCC GTC ATT CGC CTT GTT GGT      533
Cys Asp Cys Gly Gln Gly Pro Ala Leu Pro Val Ile Arg Leu Val Gly
    135                 140                 145

GGG AAC AGT GGG CAT GAA GGT CGA GTG GAG CTG TAC CAC GCT GGC CAG      581
Gly Asn Ser Gly His Glu Gly Arg Val Glu Leu Tyr His Ala Gly Gln
150                 155                 160                 165

TGG GGG ACC ATC TGT GAC GAC CAA TGG GAC AAT GCA GAC GCA GAC GTC      629
Trp Gly Thr Ile Cys Asp Asp Gln Trp Asp Asn Ala Asp Ala Asp Val
                170                 175                 180

ATC TGT AGG CAG CTG GGG CTC AGT GGC ATT GCC AAA GCA TGG CAT CAG      677
Ile Cys Arg Gln Leu Gly Leu Ser Gly Ile Ala Lys Ala Trp His Gln
            185                 190                 195
```

```
GCA CAT TTT GGG GAA GGA TCT GGC CCA ATA TTG TTG GAT GAA GTA CGC      725
Ala His Phe Gly Glu Gly Ser Gly Pro Ile Leu Leu Asp Glu Val Arg
        200                 205                 210

TGC ACC GGA AAC GAG CTG TCA ATT GAG CAA TGT CCA AAG AGT TCC TGG      773
Cys Thr Gly Asn Glu Leu Ser Ile Glu Gln Cys Pro Lys Ser Ser Trp
    215                 220                 225

GGC GAA CAT AAC TGT GGC CAT AAA GAA GAT GCT GGA GTG TCT TGT GTT      821
Gly Glu His Asn Cys Gly His Lys Glu Asp Ala Gly Val Ser Cys Val
230                 235                 240                 245

CCT CTA ACA GAT GGT GTC ATC AGA CTG GCA GGA GGA AAA AGT ACC CAT      869
Pro Leu Thr Asp Gly Val Ile Arg Leu Ala Gly Gly Lys Ser Thr His
                250                 255                 260

GAA GGT CGC CTG GAG GTC TAC TAC AAG GGG CAG TGG GGG ACA GTC TGT      917
Glu Gly Arg Leu Glu Val Tyr Tyr Lys Gly Gln Trp Gly Thr Val Cys
            265                 270                 275

GAT GAT GGC TGG ACT GAG ATG AAC ACA TAC GTG GCT TGT CGA CTG CTG      965
Asp Asp Gly Trp Thr Glu Met Asn Thr Tyr Val Ala Cys Arg Leu Leu
        280                 285                 290

GGA TTT AAA TAC GGC AAA CAG TCC TCT GTG AAC CAT TTT GAT GGC AGC      1013
Gly Phe Lys Tyr Gly Lys Gln Ser Ser Val Asn His Phe Asp Gly Ser
    295                 300                 305

AAC AGG CCC ATA TGG CTG GAT GAC GTC AGC TGC TCA GGA AAA GAA GTC      1061
Asn Arg Pro Ile Trp Leu Asp Asp Val Ser Cys Ser Gly Lys Glu Val
310                 315                 320                 325

AGC TTC ATT CAG TGT TCC AGG AGA CAG TGG GGA AGG CAT GAC TGC AGC      1109
Ser Phe Ile Gln Cys Ser Arg Arg Gln Trp Gly Arg His Asp Cys Ser
                330                 335                 340

CAT AGA GAA GAT GTG GGC CTC ACC TGC TAT CCT GAC AGC GAT GGA CAT      1157
His Arg Glu Asp Val Gly Leu Thr Cys Tyr Pro Asp Ser Asp Gly His
            345                 350                 355

AGG CTT TCT CCA GGT TTT CCC ATC AGA CTA GTG GAT GGA GAG AAT AAG      1205
Arg Leu Ser Pro Gly Phe Pro Ile Arg Leu Val Asp Gly Glu Asn Lys
        360                 365                 370

AAG GAA GGA CGA GTG GAG GTT TTT GTC AAT GGC CAA TGG GGA ACA ATC      1253
Lys Glu Gly Arg Val Glu Val Phe Val Asn Gly Gln Trp Gly Thr Ile
    375                 380                 385

TGC GAT GAC GGA TGG ACC GAT AAG CAT GCA GCT GTG ATC TGC CGG CAA      1301
Cys Asp Asp Gly Trp Thr Asp Lys His Ala Ala Val Ile Cys Arg Gln
390                 395                 400                 405

CTT GGC TAT AAG GGT CCT GCC AGA GCA AGG ACT ATG GCT TAT TTT GGG      1349
Leu Gly Tyr Lys Gly Pro Ala Arg Ala Arg Thr Met Ala Tyr Phe Gly
                410                 415                 420

GAA GGA AAA GGC CCC ATC CAC ATG GAT AAT GTG AAG TGC ACA GGA AAT      1397
Glu Gly Lys Gly Pro Ile His Met Asp Asn Val Lys Cys Thr Gly Asn
            425                 430                 435
```

```
GAG AAG GCC CTG GCT GAC TGT GTC AAA CAA GAC ATT GGA AGG CAC AAC      1445
Glu Lys Ala Leu Ala Asp Cys Val Lys Gln Asp Ile Gly Arg His Asn
        440                 445                 450

TGC CGC CAC AGT GAG GAT GCA GGA GTC ATC TGT GAC TAT TTA GAG AAG      1493
Cys Arg His Ser Glu Asp Ala Gly Val Ile Cys Asp Tyr Leu Glu Lys
    455                 460                 465

AAA GCA TCA AGT AGT GGT AAT AAA GAG ATG CTC TCA TCT GGA TGT GGA      1541
Lys Ala Ser Ser Ser Gly Asn Lys Glu Met Leu Ser Ser Gly Cys Gly
470                 475                 480                 485

CTG AGG TTA CTG CAC CGT CGG CAG AAA CGG ATC ATT GGT GGG AAC AAT      1589
Leu Arg Leu Leu His Arg Arg Gln Lys Arg Ile Ile Gly Gly Asn Asn
                490                 495                 500

TCT TTA AGG GGT GCC TGG CCT TGG CAG GCT TCC CTC AGG CTG AGG TCG      1637
Ser Leu Arg Gly Ala Trp Pro Trp Gln Ala Ser Leu Arg Leu Arg Ser
            505                 510                 515

GCC CAT GGA GAC GGC AGG CTG CTT TGT GGA GCT ACC CTT CTG AGT AGC      1685
Ala His Gly Asp Gly Arg Leu Leu Cys Gly Ala Thr Leu Leu Ser Ser
        520                 525                 530

TGC TGG GTC CTG ACA GCT GCA CAC TGC TTC AAA AGG TAC GGA AAC AAC      1733
Cys Trp Val Leu Thr Ala Ala His Cys Phe Lys Arg Tyr Gly Asn Asn
    535                 540                 545

TCG AGG AGC TAT GCA GTT CGA GTT GGG GAT TAT CAT ACT CTG GTC CCA      1781
Ser Arg Ser Tyr Ala Val Arg Val Gly Asp Tyr His Thr Leu Val Pro
550                 555                 560                 565

GAG GAG TTT GAA CAA GAA ATA GGG GTT CAA CAG ATT GTG ATT CAC AGG      1829
Glu Glu Phe Glu Gln Glu Ile Gly Val Gln Gln Ile Val Ile His Arg
                570                 575                 580

AAC TAC AGG CCA GAC AGA AGC GAC TAT GAC ATT GCC CTG GTT AGA TTG      1877
Asn Tyr Arg Pro Asp Arg Ser Asp Tyr Asp Ile Ala Leu Val Arg Leu
            585                 590                 595

CAA GGA CCA GGG GAG CAA TGT GCC AGA CTA AGC ACC CAC GTT TTG CCA      1925
Gln Gly Pro Gly Glu Gln Cys Ala Arg Leu Ser Thr His Val Leu Pro
        600                 605                 610

GCC TGT TTA CCT CTA TGG AGA GAG AGG CCA CAG AAA ACA GCC TCC AAC      1973
Ala Cys Leu Pro Leu Trp Arg Glu Arg Pro Gln Lys Thr Ala Ser Asn
    615                 620                 625

TGT CAC ATA ACA GGA TGG GGA GAC ACA GGT CGT GCC TAC TCA AGA ACT      2021
Cys His Ile Thr Gly Trp Gly Asp Thr Gly Arg Ala Tyr Ser Arg Thr
630                 635                 640                 645

CTA CAA CAA GCT GCT GTG CCT CTG TTA CCC AAG AGG TTT TGT AAA GAG      2069
Leu Gln Gln Ala Ala Val Pro Leu Leu Pro Lys Arg Phe Cys Lys Glu
                650                 655                 660

AGG TAC AAG GGA CTA TTT ACT GGG AGA ATG CTC TGT GCT GGG AAC CTC      2117
Arg Tyr Lys Gly Leu Phe Thr Gly Arg Met Leu Cys Ala Gly Asn Leu
            665                 670                 675
```

```
CAA GAA GAC AAC CGT GTG GAC AGC TGC CAG GGA GAC AGT GGA GGA CCA      2165
Gln Glu Asp Asn Arg Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro
        680                 685                 690

CTC ATG TGT GAA AAG CCT GAT GAG TCC TGG GTT GTG TAT GGG GTG ACT      2213
Leu Met Cys Glu Lys Pro Asp Glu Ser Trp Val Val Tyr Gly Val Thr
    695                 700                 705

TCC TGG GGG TAT GGA TGT GGA GTC AAA GAC ACT CCT GGA GTT TAT ACC      2261
Ser Trp Gly Tyr Gly Cys Gly Val Lys Asp Thr Pro Gly Val Tyr Thr
710                 715                 720                 725

AGA GTC CCC GCT TTT GTA CCT TGG ATA AAA AGT GTC ACC AGT CTG          2306
Arg Val Pro Ala Phe Val Pro Trp Ile Lys Ser Val Thr Ser Leu
                730                 735                 740

TAACTTATGG AAAGCTCAAG AAATAGTAAA ACAGTAACTA TTCAGTCTTC AAAAAAAAAA    2366

AAAAAAAAAA                                                           2376
```

Patent claims

1. Neurotrypsins of the formulas I and II

I: neurotrypsin of the human

II: neurotrypsin of the mouse

comprising the separate, coding and coded sequences of these compounds of the formulas I or II, comprising the separate partial sequences of the coding and coded sequences of these compounds of the formulas I or II, as for example the coding and coded sequences of the catalytic domains of the compounds of the formulas I or II, comprising the coding or coded sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the coding or coded sequences or partial sequences of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coding or coded sequences, or parts thereof, of the compounds of the formulas I or II, whose biological activity is equal or similar to that of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequence of the formulas I or II, comprising the sequences hybridizing to the coding sequences, or parts thereof, under stringent conditions, comprising the translation products of the sequences hybridizing to the coding sequences of the compounds of the formulas I or II, or to parts thereof, under stringent conditions, comprising the nucleotide sequences coding the proteins coded by the compounds of the formulas I or II, or parts thereof, but, as a result of the use of different alternative codons, are degenerated with regard to the nucleotide sequences defined by the compounds of the formulas I or II.

2. Pharmaceutical composition, characterized in that it contains as at least one active compound either the coded sequence or the coding sequence of the compound of the formula I or of the formula II, or the separate partial sequences of the coded and coding sequences of these compounds of the formulas I or II, as for example the coding or coded sequences of the catalytic domains, comprising the coding or coded sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the coding or coded sequences or partial sequences of the corresponding alleles of the compounds of the formulas I or II, comprising all

sequence variants of the coding or coded sequences, or parts th r of, of the compounds of formulas I or II, whose biological activity is qual or similar to that of the compounds of the formulas I or II, for example sequenc variants of the compounds of th formulas I or II, which differ in the not conserved amino acid sequence positions of the sequence of the formulas I or II, comprising the sequences hybridizing to the coding sequences, or parts thereof, under stringent conditions, comprising the translation products of the sequences hybridizing to the coding sequences of the compounds of the formulas I or II, or to parts thereof, under stringent conditions, comprising the nucleotide sequences coding the proteins coded by the compounds of formulas I or II, or parts thereof, but, as a result of the use of different alternative codons, are degenerated with regard to the nucleotide sequences defined by the compounds of the formulas I or II.

3. Pharmaceutical composition, characterized in that it contains as at least one active compound a substance which changes the function of the coded sequence of the compounds of formulas I or II, for example, in that it reduces or increases the catalytic activity of the coded protein, or a part thereof, or in that it shortens or prolongs the time of presence of the coded protein at its place of action in the body.

4. Pharmaceutical composition, characterized in that it contains as at least one active compound a substance which changes the expression of the coding or coded sequences of the compounds of formulas I or II, for example in that it enhances or inhibits the transcription of the mRNA or in that it enhances or inhibits the translation of the coded sequences of the compounds of formulas I or II.

5. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it prevents or reduces the growth, the expansion, the infiltration and the metastasis of primary and metastatic tumors, as for example brain tumors or tumors of the retina.

6. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it contributes to the minimization of the tissue destruction in stroke, including brain infarction and ischemia, intracerebral hemorrhage, and subarrachnoid hemorrhage, as for example by exerting a protecting effect on the cells of the so-called penumbra zone surrounding the necrotic tissue.

7. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it contributes to the minimization of the tissue destruction in traumatic brain injury, as for example by exerting a protective effect on the cells of the so-called zone surrounding the necrotic tissue.

8. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it prevents, ameliorates or cures the negative effects caused by neurodegenerative diseases.

9. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it prevents, ameliorates or cures the negative effects caused by neuroinflammatory diseases, as for example multiple sclerosis.

10. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it reduces or prevents negative effects on brain tissue caused by epileptic seizures.

11. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it contributes to the rescue of endangered neurons, as for example neurons endangered by hypoxia and ischemia, axotomy, nerve transection, deafferentiation, excitotoxicity, neuroinflammatory diseases and processes, epileptic seizures, and cancerous neoformations.

12. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it contributes to axonal regeneration and/or restoration of synaptic integrity and functions.

13. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it prevents, ameliorates, or cures retinal disorders, as for example retinal degeneration and retinal neoangiogenesis.

14. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it prevents cell death, comprising apoptosis and other forms of cell death, in the nervous system.

15. Pharmaceutical composition according to claim 14, characterized in that the cell death is an cell death in connection with damages of the nervous tissu , for xample infarct of the brain and ischemic stroke, or hemorrhage of the brain, or trauma of the brain.

16. Pharmaceutical composition according to claim 14, characterized in that the cell deathis an cell death in connection with damages of the nervous tissue, which occur due to lack of oxygen or glucose or due to intoxication.

17. Pharmaceutical composition according to claim 14, characterized in that the cell death is an cell death in connection with epileptic seizures.

18. Pharmaceutical composition according to claim 14, characterized in that the cell death is an cell death in connection with neurodegenerative diseases and inherited genetic deficiencies of the nervous system.

19. Pharmaceutical composition according to claim 14, characterized in that the cell death is an cell death in connection with axotomy or nerve transection, or deafferentiation.

20. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it influences the regeneration of injured, damaged, underdeveloped, or maldeveloped brain tissue and/or nervous tissue.

21. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it enhances the reorganization of the brain or nervous areas that have remained intact after brain and/or nerve injuries or after the destruction or damage of brain areas.

22. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it prevents, ameliorates, or cures pathological pain syndromes.

23. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it contributes to the improvement of the brain performance in healthy persons, as well as in persons with reduced brain performance.

24. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it ameliorates the learning and memory functions in healthy persons, as well as in persons with reduced learning and memory functions.

25. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it ameliorates or cures disorders in the field of disorders of the psychic wellness, or the psychosomatic state of health, as for example nervosity or „inner unrest".

26. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it ameliorates or cures disorders in the field of the emotional functions, as for example states of anxiety.

27. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it ameliorates or cures psychiatric disorders.

28. Pharmaceutical composition according to claim 27, characterized in that the psychiatric disorder is a disorder in the field of schizophrenia and schizophrenia-like disorders, comprising chronic schizophrenia, chronic schizo-affective disorders, unspecific disorders, comprising acute and chronic schizophrenia of various symptomatologies, as for example severe, non-remitting „Kraepelinic" schizophrenia, or as for example the DSM-III-R-prototype of the schizophrenia-like disorders, comprising episodic schizophrenic disorders, comprising delusionic schizophrenia-like disorders, comprising schizophrenia-like personality disorders, as for example schizophrenia-like personality disorders with mild symptomatics, comprising schizotypic personality disorders, comprising the latent forms of schizophrenic or schizophrenia-like disorders, comprising non-organic psychotic disorders.

29. Pharmaceutical composition according to claim 27, characterized in that the psychiatric disorder is a disorder in the field of the endogenic depressions or in the field of manic or manic-depressive disorders.

30. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it ameliorates or cures disorders of the brain function due deficiency, malfunction, or overfunction of at least one protease.

31. Pharmaceutical composition according to claim 30, charact rized in that the protease is tissue-type plasminogen activator, abbr viated as tPA, urokinase-type plasminogen activator, abbreviated as uPA, or plasmin.

32. Pharmaceutical composition according to claim 2, 3, or 4, characterized in that it ameliorates or cures disorders of the function of the lungs due to deficiency, malfunction, or overfunction of at least one protease.

33. Pharmaceutical composition according to claim 32, characterized in that the disorder of the function of the lungs is chronic bronchitis or emphysema of the lungs.

34. Use for the production of recombinant proteins of the coding nucleotide sequences of the compounds of the formulas I or II, comprising the separate partial sequences of the coding sequences of the compounds of the formulas I or II, as for example the coding sequences of the catalytic domains of the compounds of the formulas I or II, comprising the coding nucleotide sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the coding sequences or partial sequences thereof of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coding sequences, or parts thereof, of the compounds of formulas I or II, whose translation products have a biological activity equal or similar to that of the translation products of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequence of the formulas I or II, comprising the sequences hybridizing to the coding sequences of the compounds of the formulas I or II, or parts thereof, under stringent conditions, comprising the nucleotide sequences coding the proteins coded by the compounds of the formulas I or II, or parts thereof, but, as a result of the use of different alternative codons, are degenerated with regard to the nucleotide sequences defined by the compounds of the formulas I or II.

35. Use as targets for the development of pharmaceutical drugs, for example for the inhibition or the enhancement of the catalytic activity of the coded proteins of the formulas I or II, of proteins with the coded amino acid sequences of the compounds of the formulas I or II, comprising th proteins with the separate partial sequences of th coded amino acid sequences of the compounds of the formulas I or II, as for example

the separate catalytic domains of th compounds of the formulas I or II, comprising the proteins with the coded sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising th prot ins with the coded amino acid sequences or partial sequences thereof of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coded sequences, or parts thereof, of the compounds of formulas I or II, whose biological activity is equal or similar to the coded sequences of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequences of the formulas I or II, comprising the proteins with the coded amino acid sequences, or partial sequences thereof, of the nucleotide sequences hybridizing to the coding sequences of the compounds of the formulas I or II, or parts thereof, under stringent conditions.

36. Use as targets for the development of pharmaceutical drugs, for example for the enhancement or the inhibition of the catalytic activity of the coded proteins of the formulas I or II, of the species-homologous proteins, or parts thereof, of the compounds of the formulas I or II, as for example the species-homologous proteins of the rat, the rabbit, the cow, the sheep, the pig, the primates, the birds, the zebra fish, the fruit fly (Drosophila melanogaster), etc., comprising the partial sequences thereof, as for example the separate catalytic domains, comprising the splice variants of the species-homologous proteins, comprising the alleles of the species-homologous proteins, comprising the translation products of the sequences hybridizing under stringent conditions to the corresponding species-homologous compounds of the formulas I or II, or their splice variants, or their alleles, of the coding sequences or partial sequences of the compounds of formulas I or II .

37. Use for the spatial structure determination, for example the spatial structure determination by means of crystallography or nuclear resonance spectroscopy, of the proteins with the coded amino acid sequences of the compounds of the formulas I or II, comprising the proteins with the separate partial sequences of the coded amino acid sequences of the compounds of the formulas I or II, as for example the separate catalytic domains, comprising the proteins with the coded sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising th prot ins with the coded amino acid sequences, or partial sequences th reof, of the corresponding alleles of the compounds of the formulas I or II, comprising

all sequence variants of the coded sequences, or parts thereof, of the compounds of the formulas I or II, whose biological activity is equal or similar to that of the coded sequences of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequences of the formulas I or II, comprising the translation products with the sequences hybridizing to the coding sequences of the compounds of the formulas I or II, or parts thereof, under stringent conditions, comprising the species-homologous proteins of the compounds of the formulas I or II, for example the species-homologous proteins of the rat, the rabbit, the cow, the sheep, the pig, the primates, the birds, the zebra fish, the fruit fly (Drosophila melanogaster), etc., comprising the partial sequences thereof, as for example the separate catalytic domains.

38. Use for the prediction of the protein structure by means of computerized protein structure prediction methods, of the coded amino acid sequences of the compounds of the formulas I or II, comprising the separate partial sequences of the coded amino acid sequences of the compounds of the formulas I or II, as for example the coded amino acid sequences of the separate catalytic domains of the compounds of the formulas I or II, comprising the coded sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the coded amino acid sequences, or parts thereof, of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coded sequences, or parts thereof, of the compounds of the formulas I or II, whose biological activity is equal or similar to that of the coded sequences of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequences of the formulas I or II, comprising the amino acid sequences of the translation products of the sequences hybridizing to the coding sequences of the compounds of the formulas I or II, or parts thereof, under stringent conditions, comprising sequences of the species-homologous compounds of the compounds of the formulas I or II, for example the sequences of the species-homologous compounds of the rat, the rabbit, the cow, the sheep, the pig, the primates, the birds, the zebra fish, the fruit fly (Drosophila melanogaster), etc., comprising the partial sequences of the species-homologous compounds, as for example the sequences of the catalytic domains of the species-homologous compounds.

39. Us as targets for the development of pharmaceutical drugs , for xample for the inhibition or the enhancement of the catalytic activity of the coded proteins of the compounds of the formulas I or II, of the spatial structure of th coded amino acid sequences of the compounds of the formulas I or II, comprising the spatial structures of the separate partial sequences of the compounds of the formulas I or II, as for example the spatial structure of the catalytic domains, comprising the spatial structure of the coded sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the spatial structure of the coded sequences or partial sequences of the corresponding alleles of the compounds of the formulas I or II, comprising the spatial structure of all sequence variants of the coded sequences, or parts thereof, of the compounds of formulas I or II, whose biological activity is equal or similar to the coded sequences of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequences of the formulas I or II, comprising the spatial structures of the translation products of the sequences hybridizing to the coding sequences of the compounds of the formulas I or II, or parts thereof, under stringent conditions, comprising the spatial structures of the species-homologous compounds of the compounds of the formulas I or II, as for example the spatial structures of the species homologous compounds, or parts thereof, of the rat, the rabbit, the cow, the sheep, the pig, the primates, the birds, the zebra fish, the fruit fly (Drosophila melanogaster), etc..

40. Use in gene therapeutical applications in humans and in animals, as for example as parts of gene therapy vectors or as for example as parts of artificial chromosomes, of the coding nucleotide sequences of the compounds of the formulas I or II, comprising the separate partial sequences of the coding sequences of these compounds of the formulas I or II, as for example the coding sequences of the catalytic domains of the compounds of the formulas I or II, comprising the coding sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the coding sequences or partial sequences of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coding sequences, or parts thereof, of the compounds of the formulas I or II, whose translation products exhibit a biological activity which is equal or similar to that of the translation products of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino

acid sequence positions of the sequences of the compounds of the formulas I or II, comprising the sequences hybridizing to the coding sequences, or parts thereof, under stringent conditions, comprising the nucleotide s quences coding the proteins coded by the compounds of the formulas I or II, or parts thereof, but as a result of the use of different alternative codons, are degenerated with regard to the nucleotide sequences defined by the compounds of the formulas I or II.

41. Use for so-called cell engineering applications for the production of gene technologically mutated cells, which produce the coded sequences, or parts thereof, of the compounds of the formulas I or II, for example for cell-therapeutical applications as a pharmaceutical composition according to claim 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, or 33, of the coding nucleotide sequences of the compounds of the formulas I or II, comprising the separate partial sequences of the coding sequences of these compounds of the formulas I or II, as for example the coding sequences of the catalytic domains of the compounds of the formulas I or II, comprising the coding sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the coding sequences or partial sequences of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coding sequences, or parts thereof, of the compounds of the formulas I or II, whose translation products exhibit a biological activity which is equal or similar to that of the translation products of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequence of the compounds of the formulas I or II, comprising the sequences hybridizing to the coding sequences, or parts thereof, under stringent conditions, comprising the nucleotide sequences coding the proteins coded by the compounds of formulas I or II, or parts thereof, but as a result of the use of different alternative codons, are degenerated with regard to the nucleotide sequences defined by the compounds of the formulas I or II.

42. Use as antigens for the production of antibodies, as for example antibodies that inhibit or promote the protease function or antibodies that can be used for immunohistochemical studies, of the coded amino acid sequences of the compounds of th formulas I or II, comprising the separate partial sequences of th coded amino acid sequences of the compounds of the formulas I or II, as for example the coded amino

acid sequence of the catalytic domain or one or more of the other domains or segments, comprising the cod d sequences or partial sequences of th corresponding splice variants of the compounds of the formulas I or II, comprising the coded sequences or partial sequences of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coded sequences, or parts thereof, of the compounds of the formulas I or II, whose biological activity is equal or similar to that of the coded sequences of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequence of the compounds of the formulas I or II, comprising the translation products or parts thereof, of the sequences hybridizing to the coding sequences of the compounds of the formulas I or II, or parts thereof, under stringent conditions, comprising the coded sequences of the species-homologous compounds of the compounds of the formulas I or II, as for example the coded sequences of the species-homologous compounds of the rat, the rabbit, the cow, the sheep, the pig, the primates, the birds, the zebra fish, the fruit fly (Drosophila melanogaster), etc., comprising the separate partial sequences of the coded sequences of the species-homologous compounds of the compounds of the formulas I or II, as for example the coded amino acid sequence of the catalytic domain, or one or more of the other domains or segments.

43. Use for the production of transgenic animals, as for example transgenic mice, of the coding nucleotide sequences of the compounds of the formulas I or II, comprising the separate partial sequences of the coding sequences of these compounds of the formulas I or II, as for example the coding sequences of the catalytic domains of the compounds of the formulas I or II, comprising the coding sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the coding sequences, or partial sequences, of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coding sequences, or parts thereof, of the compounds of the formulas I or II, whose translation products exhibit a biological activity which is equal or similar to that of the translation products of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequences of the compounds of the formulas I or II, comprising the sequenc s hybridizing to the coding sequences, or parts thereof, under stringent conditions, comprising the nucleotide sequences coding the prot ins coded by the

compounds of the formulas I or II, or parts thereof, but as a result of the use of different alternative codons, are degenerated with regard to the nucleotide sequences defined by the compounds of the formulas I or II.

44. Use for the inactivation or the mutation of the corresponding gene by means of gene targeting techniques, as for example the elimination of the gene in the mouse through homologous recombination, of the coding nucleotide sequences of the compounds of the formulas I or II, comprising the separate partial sequences of the coding sequences of these compounds of the formulas I or II, as for example the coding sequences of the catalytic domains of the compounds of the formulas I or II, comprising the coding sequences, or partial sequences, of the corresponding splice variants of the compounds of the formulas I or II, comprising the coding sequences, or partial sequences, of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coding sequences, or parts thereof, of the compounds of the formulas I or II, whose translation products exhibit a biological activity which is equal or similar to that of the translation products of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequence of the compounds of the formulas I or II, comprising the sequences hybridizing to the coding sequences, or parts thereof, under stringent conditions, comprising the nucleotide sequences coding the proteins coded by the compounds of the formulas I or II, or parts thereof, but as a result of the use of different alternative codons, are degenerated with regard to the nucleotide sequences defined by the compounds of the formulas I or II.

45. Use for the diagnostics of disorders in the gene corresponding to the compound of the formula I, of the coding nucleotide sequences of the compounds of the formulas I or II, comprising the separate partial sequences of the coding sequences of these compounds of the formulas I or II, as for example the coding sequences of the catalytic domains of the compounds of the formulas I or II, comprising the coding sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the coding sequences, or partial sequences, of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coding sequences, or parts thereof, of the compounds of the formulas I or II, whose translation products exhibit a biological activity which is equal or similar to that of the translation products of the compounds of the formulas I or II, for example

sequence variants of the compounds of th formulas I or II, which differ in the not conserved amino acid sequence positions of the sequenc s of the compounds of the formulas I or II, comprising the sequences hybridizing to th coding sequences, or parts thereof, under stringent conditions, comprising the nucleotide sequences coding the proteins coded by the compounds of the formulas I or II, or parts thereof, but as a result of the use of different alternative codons, are degenerated with regard to the nucleotide sequences defined by the compounds of the formulas I or II.

46. Use as a starting sequence for gene technological modifications aimed at the production of pharmaceutical compositions or gene therapy vectors which exhibit changed properties as compared with the corresponding pharmaceutical compositions or gene therapy vectors containing the coding nucleotide sequence of the compounds of formulas I or II, for example changed proteolytic activity, changed proteolytic specificity, or changed pharmacokinetic characteristics, of the coding nucleotide sequences of the compounds of the formulas I or II, comprising the separate partial sequences of the coding sequences of these compounds of the formulas I or II, as for example the coding sequences of the catalytic domains of the compounds of the formulas I or II, comprising the coding sequences or partial sequences of the corresponding splice variants of the compounds of the formulas I or II, comprising the coding sequences, or partial sequences, of the corresponding alleles of the compounds of the formulas I or II, comprising all sequence variants of the coding sequences, or parts thereof, of the compounds of the formulas I or II, whose translation products exhibit a biological activity which is equal or similar to that of the translation products of the compounds of the formulas I or II, for example sequence variants of the compounds of the formulas I or II, which differ in the not conserved amino acid sequence positions of the sequences of the compounds of the formulas I or II, comprising the sequences hybridizing to the coding sequences, or parts thereof, under stringent conditions, comprising the nucleotide sequences coding the proteins coded by the compounds of the formulas I or II, or parts thereof, but as a result of the use of different alternative codons, are degenerated with regard to the nucleotide sequences defined by the compounds of the formulas I or II.

INTERNATIONAL SEARCH REPORT

national Application No
PCT/IB 98/00625

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/57 C12N9/64 A61K31/70 A61K38/48 C12Q1/37
A61K48/00 C07K16/40 C12N15/00 A01K67/027 C12Q1/68

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N A61K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL/GENBANK DATABASES Accesion no AA166524, Sequence identification MMAA66524, 21 December 1996 MARRA M ET AL:"THe WashU-HHMI Mouse EST Project" XP002075099 see the whole document	1,34,37, 38,42
X	EMBL/GENBANK DATABASES Accesion no AA373034, Sequence identification HSZZ78160, 18 April 1997 ADAMS M ET AL: "Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence" XP002075100 see the whole document	1,34,37, 38,42

-/--

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 August 1998	04/09/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Van der Schaal, C

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 98/00625

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL/GENBANK DATABASES Accesion no AA073513, Sequence identification MMA73513, 5 October 1996 MARRA M ET AL:"THe WashU-HHMI Mouse EST Project" XP002075101 see the whole document	1,34,37, 38,42
X	EMBL/GENBANK DATABASES Accesion no AA063841, Sequence identification MMA63841, 25 September 1996 MARRA M ET AL:"THe WashU-HHMI Mouse EST Project" XP002075102 see the whole document	1,34,37, 38,42
A	YAMASHIRO K ET AL: "Molecular cloning of a novel trypsin-like serine protease (neurosin) preferentially expressed in brain." BIOCHIMICA ET BIOPHYSICA ACTA, (1997 JAN 3) 1350 (1) 11-4. JOURNAL CODE: AOW. ISSN: 0006-3002., XP002075096 Netherlands	
P,X	DATABASE MEDLINE US NATIONAL LIBRARY OF MEDICINE (NLM), BETHESDA, MD, US PROBA K ET AL: "Cloning and sequencing of the cDNA encoding human neurotrypsin." XP002075103 see abstract & BIOCHIMICA ET BIOPHYSICA ACTA, (1998 MAR 9) 1396 (2) 143-7. JOURNAL CODE: AOW. ISSN: 0006-3002., Netherlands	1-4, 34-39, 42-44
P,X	GSCHWEND T P ET AL: "Neurotrypsin, a novel multidomain serine protease expressed in the nervous system." MOLECULAR AND CELLULAR NEUROSCIENCES, (1997) 9 (3) 207-19. JOURNAL CODE: B1D. ISSN: 1044-7431., XP002075097 United States see the whole document	1-46

-/—

3

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

I. .ational Application No
PCT/IB 98/00625

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	YAMAMURA Y ET AL: "Molecular cloning of a novel brain-specific serine protease with a kringle-like structure and three scavenger receptor cysteine-rich motifs." BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, (1997 OCT 20) 239 (2) 386-92. JOURNAL CODE: 9Y8. ISSN: 0006-291X., XP002075098 United States see the whole document	1-4, 34-39, 42-44

3

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 98/00625

Box I Observations where certain claims were found uns archabl (C ntinuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: Although claims 40 42 44 and 45 are (partially) directed to a method of treatment of the human/animal body or to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.